DOCUMENT RESUME

ED 201 644 TM 810 187

AUTHOR Fortney, William G.; Miller, Robert B.
TITLE Bayesian Analysis in Random Coefficient m-group Regression.
PUB DATE [80]
NOTE 22p.; Address correspondence to Robert B. Miller, Department of Statistics and Graduate School of Business, University of Wisconsin, Madison, WI 53706.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bayesian Statistics; *Least Squares Statistics; *Mathematical Models; *Multiple Regression Analysis
IDENTIFIERS Estimation (Mathematics); *Monte Carlo Studies

ABSTRACT

Bayesian analysis of an m-group model is considered. A convenient stage III prior is proposed, and cases when the posterior distributions take on a simple form are exhibited. The behavior of various point estimators of the linear parameters of the model are explored in a Monte Carlo study. In the simple model considered, the O'Hagan estimator usually displays smaller mean squared error than the competing estimators. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201644

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

BAYESIAN ANALYSIS IN RANDOM COEFFICIENT m-group REGRESSION

William G. Fortney
Boeing Computer Services Co.

and

Robert B. Miller[+]
Department of Statistics
and
Graduate School of Business

University of Wisconsin, Madison
1210 W. Dayton Street
Madison, WI 53706

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
R.B. Miller
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 810 187

[+]To whom correspondence should be addressed.

Bayesian Analysis in Random Coefficient m-group Regression

by

William G. Fortney
Robert B. Miller

ABSTRACT

Bayesian analysis of an m-group regression model is considered. A convenient stage III prior is proposed, and cases when the posterior distributions take on a simple form are exhibited. The behavior of various point estimators of the linear parameters of the model are explored in a Monte Carlo study.

Key Words: BAYESIAN, RANDOM COEFFICIENTS, M-GROUP REGRESSION, MULTIVARIATE MATRICVARIATE T, JOINT MODAL ESTIMATOR, MARGINAL MODAL ESTIMATOR, O'HAGAN ESTIMATOR, MONTE CARLO STUDY.

1. INTRODUCTION

Lindley and Smith (1972) proposed a hierarchical m-group regression model consisting of three stages. Whereas we adopt their stages I and II intact, we propose a modification of their stage III. The purpose of the modification is to specify prior distributions that assign probability zero to singular stage II covariance matrices. This is a useful modification because other proposed stage III priors have permitted the assumption of singular stage II covariance matrices in such a way that it could not be overcome by data, thus producing bimodal posterior distributions. See Fortney and Miller (1980a), (1980b) for a discussion of this phenomenon in the hierarchical multiple regression model.

A simplifying linear transformation permits us to see when the analysis of the m-group regression model is equivalent to the analysis of a single multiple regression model.

Finally, a small Monte Carlo experiment compares various estimators that have been proposed for the stage I parameters of the m-group model.

2. THE m-group REGRESSION MODEL

2.1 Assumptions

In our notation the m-group regression model is as follows.

Stage I

$$\underset{\sim}{y}_i = \underset{\sim}{X}_i\underset{\sim}{\beta}_i + \underset{\sim}{\varepsilon}_i, \quad i = 1,\ldots,m,$$

where $\underline{y}_i' = (y_{i1},\ldots,y_{in_i})$ is an $n_i \times 1$ vector of observations, $\underline{X}_i = ((x_{jk}^{(i)}))$ is an $n_i \times p$ matrix of known constants, $\underline{\beta}_i' = (\beta_{i1},\ldots,\beta_{ip})$ is a $p \times 1$ vector of unknown parameters, and $\underline{\varepsilon}_i' = (\varepsilon_{i1},\ldots,\varepsilon_{in_i})$ is an $n_i \times 1$ stochastic vector whose components are $NID(0,\sigma^2)$. We further assume $\underline{\varepsilon}_1,\ldots,\underline{\varepsilon}_m$ are stochastically independent.

Stage II

$$\underline{\beta}_1,\ldots,\underline{\beta}_m \text{ are } NID_p(\underline{\xi},\underline{\Sigma}).$$

Stage III

Assume $\underline{\xi}$, σ^2, and $\underline{\Sigma}$ are stochastically independent *a priori*, a flat (improper) prior on $\underline{\xi}$, and an inverted gamma prior on σ^2 with density proportional to

$$(\sigma^2)^{-(\nu+2)/2} \exp\{-\nu\lambda/2\sigma^2\}. \tag{1}$$

The mode of this density is $m_0 = \nu\lambda/(\nu+2)$, the mean is $\nu\lambda/(\nu-2)$, provided $\nu > 2$, and the variance is $2\nu^2\lambda^2/(\nu-2)^2(\nu-4)$, provided $\nu > 4$. The density is not infinite-valued for any positive value of ν, and is zero when $\sigma^2 = 0$ (provided $\nu > 0$, of course). Parameterized in terms of ν and m_0, the density in (1) is proportional to

$$(\sigma^2)^{-(\nu+2)/2} \exp\{-(\nu+2)m_0/2\sigma^2\}. \tag{2}$$

If we let $\nu \downarrow 0$ in this density, we get the improper density $\sigma^{-2}\exp\{-m_0/\sigma^2\}$, which approaches zero as $\sigma^2 \downarrow 0$. The rationale for this parameterization is that an investigator may be able to specify a modal value for σ^2 *a priori* but without much precision. Intuitively, this yields the limiting process described above. Note that had we allowed $\nu \downarrow 0$ for fixed λ in (1), then (1) would have approached the improper density proportional to σ^{-2}, which is infinite-valued at $\sigma^2 = 0$. Such meaningless prior

specifications are not necessarily overcome in hierarchical models because the data may contain only weak information about stage II parameters. (See also Goel and DeGroot (1979).) Thus we shall adopt the density in (2) as our prior for σ^2, permitting $\nu = 0$ as a possibility.

We shall assume that a priori $\underline{\Sigma}$ has the inverted Wishart distribution, denoted by $W^{-1}(\underline{R},\eta+p+1,p)$, with density proportional to

$$|\underline{\Sigma}|^{-(\eta+p+1)/2} \exp\{-\tfrac{1}{2}\mathrm{tr}(\underline{R}\underline{\Sigma}^{-1})\}. \tag{3}$$

The mode of this density is $\underline{M}_0 = \underline{R}/(\eta+p+1)$, and the mean is $\underline{R}/(\eta-p-1)$, provided $\eta > p + 1$. For $\eta > p - 1$, the density is proper, the mode exists, and the density is zero if $\underline{\Sigma}$ is non-negative definite and singular. Parameterized in terms of η and $\underline{M}_0$, the density in (3) is proportional to

$$|\underline{\Sigma}|^{-(\eta+p+1)/2} \exp\{-[(\eta+p+1)/2]\mathrm{tr}\underline{M}_0\underline{\Sigma}^{-1}\}. \tag{4}$$

Its limit (for fixed $\underline{M}_0$) as $\eta\downarrow(p-1)$ is $|\underline{\Sigma}|^{-p} \exp\{-p\mathrm{tr}\underline{M}_0\underline{\Sigma}^{-1}\}$, an improper density that is zero-valued for singular $\underline{\Sigma}$ and has mode $\underline{M}_0$. The argument for choosing (4) is analogous to that for choosing (2). However, Lindley and Smith (1972) chose the prior assumption that $\underline{\Sigma}^{-1}$ was distributed as $W(\underline{R}^{-1},\eta,p)$. The resulting density was proper for $\eta > p - 1$, but its mode was finite only if $\eta > p + 1$. Moreover, for $p - 1 < \eta \leq p + 1$, the density was infinite-valued for $\underline{\Sigma}^{-1}$ non-negative definite and singular. (For a graph of this situation see Fortney and Grupe (1976).) The derivations of the modes of the two $\underline{\Sigma}$ densities discussed here are presented in Appendix A.

2.2 Posterior Distributions

Define the $m \times p$ matrix $\underline{B} = (\underline{\beta}_1,\ldots,\underline{\beta}_m)'$. The assumptions of section 2.1 easily lead to the joint posterior distribution of $\underline{B}$, σ^2, $\underline{\Sigma}$, and $\underline{\xi}$, given $\underline{y}_1,\ldots,\underline{y}_m$. We treat $\underline{\xi}$ as a nuisance parameter and integrate it out to find the joint posterior density of $\underline{B}$, σ^2, and $\underline{\Sigma}$ proportional to the product of the two factors

$$(\sigma^2)^{-(\nu+n+2)/2} \exp\{-(2\sigma^2)^{-1}[r+\sum_{i=1}^{m}(\underline{\beta}_i-\hat{\underline{\beta}}_i)'\underline{X}_i'\underline{X}_i(\underline{\beta}_i-\hat{\underline{\beta}}_i)]\} \tag{5a}$$

and

$$|\underline{\Sigma}|^{-(\eta+m+p)/2} \exp\{-\tfrac{1}{2}\mathrm{tr}\underline{\Sigma}^{-1}[(\eta+p+1)\underline{M}_o + \sum_{i=1}^{m}(\underline{\beta}_i-\bar{\underline{\beta}})(\underline{\beta}_i-\bar{\underline{\beta}})']\} \tag{5b}$$

where $n = n_1+\ldots+n_m$, $\bar{\underline{\beta}} = m^{-1}\sum_{i=1}^{m}\underline{\beta}_i$, $\hat{\underline{\beta}}_i$ is the ordinary least squares estimator of $\underline{\beta}_i$ using the equation of the i-th group only, and $r = (\nu+2)m_o + \sum_{i=1}^{m}(\underline{y}_i-\underline{X}_i\hat{\underline{\beta}}_i)'(\underline{y}_i-\underline{X}_i\hat{\underline{\beta}}_i)$.

Integrating σ^2 and $\underline{\Sigma}$ out of this density, we get the marginal posterior density of $\underline{B}$ proportional to the product of the two factors

$$[r+\sum_{i=1}^{m}(\underline{\beta}_i-\hat{\underline{\beta}}_i)'\underline{X}_i'\underline{X}_i(\underline{\beta}_i-\hat{\underline{\beta}}_i)]^{-(\nu+n)/2} \tag{6a}$$

and

$$|(\eta+p+1)\underline{M}_o + \sum_{i=1}^{m}(\underline{\beta}_i-\bar{\underline{\beta}})(\underline{\beta}_i-\bar{\underline{\beta}})'|^{-(\eta+m+1)/2}. \tag{6b}$$

The factor in (6a) is proportional to a multivariate-t density. The factor in (6b) can be written

$$|(\eta+p+1)\underline{M}_o + \underline{B}'\underline{Q}_m\underline{B}|^{-(\eta+m+1)/2} \tag{6b'}$$

where the (i,j)-th element of $\underline{Q}_m$ is $\delta_{ij} - m^{-1}$, and δ_{ij} is the Kronecker δ.

Thus (6b) is proportional to a matricvariate-t density. We shall refer to a density that is proportional to the product of multivariate- and matricvariate-t densities as MMT.

2.3 Reductions to Multiple Regression Analysis

In another paper, Fortney and Miller (1980b), we show how to obtain first-and second-order moments of densities that are products of multivariate-t densities by making a linear transformation to "canonical form". A similar transformation will now be exhibited for the MMT density. This transformation will allow us to see when the analysis of the MMT density reduces to the analysis of the product of multivariate-t densities.

First select an orthogonal m×m matrix $H = ((h_{ij}))$ such that $\underline{Q}_m = H\underline{\Lambda}_m H'$, $\underline{\Lambda}_m = \text{diag}(0,1,\ldots,1)$. Also choose a p×p non-singular matrix $\underline{U}$ such that $(\eta+p+1)\underline{M}_o = \underline{U}'\underline{U}$. Then define the random matrix $\underline{\Gamma} = ((\gamma_{ij}))$ by

$$\underline{B} = m^{-1/2}\underline{H}\underline{\Gamma}\underline{U} \longleftrightarrow \underline{\Gamma} = m^{1/2}\underline{H}'\underline{B}\underline{U}^{-1} \tag{7}$$

The Jacobian of this transformation is $m^{-mp/2}|(\eta+p+1)\underline{M}_o|^{m/2}$. Under this transformation the factor in (6a) is proportional to

$$\{1+(\underline{\gamma}-\hat{\underline{\gamma}})'[\nu_1^{-1}\underline{S}](\underline{\gamma}-\hat{\underline{\gamma}})\}^{-(\nu+n)/2} \tag{8}$$

where $\nu_1 = \nu + n - mp$, $\underline{S}$ is an mp×mp matrix consisting of m^2 blocks of size p×p, of which the block in the (i,j)-th position is

$$\underline{S}_{ij} = (\nu_1/mr)\underline{U}\left(\sum_{k=1}^{m} h_{ki}h_{kj}\underline{X}_k'\underline{X}_k\right)\underline{U}' \tag{9}$$

where $\hat{\underline{\Gamma}} = ((\hat{\gamma}_{ij}))$ is defined by $\hat{\underline{B}} = m^{-1/2}\underline{H}\hat{\underline{\Gamma}}\underline{U}$, $\underline{\gamma} = P(\underline{\Gamma}')$, $\hat{\underline{\gamma}} = P(\hat{\underline{\Gamma}}')$, and $P(\cdot)$ is the pack operator. (See Appendix B for details.) The

factor in (6b) is easily seen to be proportional to

$$|\underline{I}+m^{-1}\underline{\Gamma}'\underline{\Lambda}_m\underline{\Gamma}|^{-(\eta+m-1)/2} . \qquad (10)$$

The factor in (8) is proportional to a multivariate-t density. Cases where (10) is also proportional to a multivariate-t density are of special interest because the techniques we developed for obtaining moments of $\underline{\Gamma}$ (and hence of $\underline{B}$) are immediately applicable. Two cases are worthy of mention. The first is when $p = 1$, for then $\underline{\Gamma}$ is $m\times 1$, and $|\underline{I}+m^{-1}\underline{\Gamma}'\underline{\Lambda}_m\underline{\Gamma}| = 1+m^{-1}\sum_{i=2}^{m}\gamma_{i1}^2$. This model would arise if (a) the means of m populations were to be estimated, and the means were assumed to be independently drawn from a normal population, or (b) each of the m regression curves were a straight line through the origin.

The second case is when $m = 2$. The singularity of $\underline{Q}_m$ yields $\underline{\Gamma}\underline{\Lambda}_m\underline{\Gamma} = \gamma_2\gamma_2'$ and $|\underline{I}+m^{-1}\underline{\Gamma}'\underline{\Lambda}_m\underline{\Gamma}| = |\underline{I}+m^{-1}\gamma_2\gamma_2'| = 1 + m^{-1}\gamma_2\gamma_2$, so that the factor in (10) is again proportional to a multivariate-t density.

Except for the cases when the MMT reduces to a product of multivariate densities, there appears to be no economical means of calculating the moments of the MMT. A very complicated asymptotic expansion generalizing that of Tiao and Zellner (1964) is possible, but it is not useful for the MMT that arises in the hierarchical model. The reason is that the distribution is not asymptotically normal. Whereas the first factor (5a) approaches normality with increasing sample size, the second factor (5b) does not.

3. ESTIMATORS

The only type of inference considered in the literature for the MMT model so far has been the production of point estimators. In view of the difficulties encountered in studying other aspects of the MMT, it is important to try to appreciate the reliability of these estimators. In this section we display formulas, in our notation, for three point estimators, and in the next section we study their behavior via a Monte Carlo experiment. Since our simulations assume a common $\underset{\sim}{X}$ matrix for all regressions, we quote the formulas under this assumption.

Lindley and Smith (1972) suggested that point estimators of $\underset{\sim}{B}$ be found by computing the mode of the joint posterior distribution of <u>all</u> the parameters. In our formulation this requires the maximization with respect to $\underset{\sim}{B}$, σ^2, and $\underset{\sim}{\Sigma}$ of the density proportional to the product of the factors in (5a) and (5b). A set of equations that may be solved interatively for the resulting modal estimators is

$$\underset{\sim}{\beta}_i = (\underset{\sim}{X}'\underset{\sim}{X}+\sigma^2\underset{\sim}{\Sigma}^{-1})^{-1}(\underset{\sim}{X}'\underset{\sim}{y}_i+\sigma^2\underset{\sim}{\Sigma}^{-1}\bar{\underset{\sim}{\beta}}),\ \bar{\underset{\sim}{\beta}} = m^{-1}\sum_{i=1}^{m}\underset{\sim}{\beta}_i \tag{11a}$$

$$\sigma^2 = (n+\nu+2)^{-1}[r+\sum_{i=1}^{m}(\underset{\sim}{\beta}_i-\hat{\underset{\sim}{\beta}}_i)'\underset{\sim}{X}'\underset{\sim}{X}(\underset{\sim}{\beta}_i-\hat{\underset{\sim}{\beta}}_i)] \tag{11b}$$

$$\underset{\sim}{\Sigma} = (m+\eta+p)^{-1}[(\eta+p+1)\underset{\sim}{M}_o + \sum_{i=1}^{m}(\underset{\sim}{\beta}_i-\bar{\underset{\sim}{\beta}})(\underset{\sim}{\beta}_i-\bar{\underset{\sim}{\beta}})'] \ . \tag{11c}$$

The estimators of the $\underset{\sim}{\beta}_i$'s resulting from these equations will be called joint modal estimators (JME).

Another set of modal estimators may be obtained by maximizing the density proportional to the product of the two factors in

(6a) and (6b). We shall call these estimators marginal modal estimators (MME). A set of equations that may be solved iteratively for the MME is equation (11a) along with

$$\sigma^2 = (n+\nu)^{-1}[r+\sum_{i=1}^{m}(\underline{\beta}_i-\hat{\underline{\beta}}_i)'\underline{X}'\underline{X}(\underline{\beta}_i-\hat{\underline{\beta}}_i)] \tag{12}$$

and

$$\underline{\Sigma} = (m+\eta-1)^{-1}[(\eta+p+1)\underline{M}_o + \sum_{i=1}^{m}(\underline{\beta}_i-\bar{\underline{\beta}})(\underline{\beta}_i-\bar{\underline{\beta}})'] \ . \tag{13}$$

The final set of estimators we consider was suggested by O'Hagan (1976). It is the mode (also mean) of the normal conditional distribution of $\underline{B}$, given σ^2 and $\underline{\Sigma}$, where σ^2 and $\underline{\Sigma}$ are replaced by the modal values of their joint posterior distribution. We shall call these estimators O'Hagan estimators. A set of equations that can be solved iteratively for O'Hagan's estimators is equation (11a) along with

$$\sigma^2 = \frac{r+\sum_{i=1}^{m}(\underline{\beta}_i-\hat{\underline{\beta}}_i)'\underline{X}'\underline{X}(\underline{\beta}_i-\hat{\underline{\beta}}_i)}{n+\nu+2-p-(m-1)\mathrm{tr}[\underline{X}'\underline{X}(\underline{X}'\underline{X}+\sigma^2\underline{\Sigma}^{-1})^{-1}]} \tag{14}$$

and

$$\underline{\Sigma} = (m+\eta+p)^{-1}[(\eta+p+1)\underline{M}_o + \sum_{i=1}^{m}(\underline{\beta}_i-\bar{\underline{\beta}})(\underline{\beta}_i-\bar{\underline{\beta}})' + (m-1)\sigma^2(\underline{X}'\underline{X}+\sigma^2\underline{\Sigma}^{-1})^{-1}] \tag{15}$$

Whereas the assumption of common $\underline{X}$ for all groups is not essential for the calculation of JME and MME, it does appear to be necessary for the derivation of the equations leading to O'Hagan's estimators.

4. A MONTE CARLO STUDY

In this section we present the results of a small Monte Carlo experiment designed to compare the estimators outlined in section 3, assuming that the hierarchical m-group regression model is true. We incidentally compare these estimators to ordinary least squares estimates also.

4.1 Cases Simulated

Our primary concern is to study the impact of various factors on the precision of the various point estimators. The simulation environments all assumed simple linear regression (p=2). The four factors investigated were sample size (n_i), group size (m), degree of corrleation between columns of $\underset{\sim}{X}$ (denoted by r_x), and degree of correlation between elements of $\underset{\sim}{\beta}_i$ (denoted by ρ). We took

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \text{ and } \underset{\sim}{X}'\underset{\sim}{X} = \begin{bmatrix} 1 & r_x \\ r_x & 1 \end{bmatrix}$$

in all cases.

Sixteen cases were generated by taking each of the four factors at two levels, as outlined in Table 1. Each case was replicated thirty times using $\sigma^2 = .0625$ and $\underset{\sim}{\xi}' = [0,0]$. For the computation of the estimators we took $\nu = \lambda = \eta = 0$ and $\underset{\sim}{M}_0 = \underset{\sim}{\Sigma}$. In practice this lucky prior mode would seldom be attained, but we decided to evaluate the estimates under ideal conditions.

4.2 Presentation of Results

As criteria for evaluating the estimators we used the two quadratic forms

$$d_1 = m^{-1} \sum_{t=1}^{30} \sum_{i=1}^{m} (\underset{\sim}{\beta}^*_{it} - \underset{\sim}{\beta}_{it})'(\underset{\sim}{\beta}^*_{it} - \underset{\sim}{\beta}_{it})$$

and

$$d_2 = m^{-1} \sum_{t=1}^{30} \sum_{i=1}^{m} (\underset{\sim}{\beta}^*_{it} - \underset{\sim}{\beta}_{it}) \underset{\sim}{\Sigma}^{-1} (\underset{\sim}{\beta}^*_{it} - \underset{\sim}{\beta}_{t}),$$

where $\underset{\sim}{\beta}_{it}$ denotes the value of $\underset{\sim}{\beta}_i$ on the t-th replication, and $\underset{\sim}{\beta}^*_{it}$ denotes any one of the four estimators of $\underset{\sim}{\beta}_{it}$. Since the rankings of the estimators were the same for d_1 and d_2, only the d_2 criterion is reported.

Table 2 reports the results of the Monte Carlo study. The average values of d_2 (over 30 replications) are reported for the JME, MME, and O'Hagan estimators. The ratios of these averages to the average d_2 of the ordinary least squares estimator is quoted as a percentage in parentheses. The "best" estimator (lowest average d_2) receives a star. Finally, the ranks over the sixteen cases of the ratios of the best estimators are reported, with the greatest improvement receiving rank 1. O'Hagan's estimator is best in fourteen of the sixteen cases.

The trends in the ranks are suggestive. Generally, but not without exceptions, ranks decrease as we move from lesser to greater group size (m=8 to m=20), lesser to greater multicollinearity (r_x=0 to r_x=.5), and lesser to greater correlation between β'S (ρ=0 to ρ=.5). There is no systematic advantage to increasing the sample size n_i.

Table 3 shows the marginal averages of d_2 (over 240 replications) for each of the four factors. Here O'Hagan's estimate is clearly the best, and the JME are the worst, but the differences are small.

In summary, the simulation study suggests that when the m-equation model applies, the O'Hagan estimators will typically enjoy the lowest mean squared error of the estimators considered, and that all the estimators will be superior to the ordinary least squares estimator. However, it may be difficult to derive formulas for O'Hagan's estimator for some models of interest, e.g. for the model discussed by Novick, et. al (1972).

4.3 A Remark on Computation

There were no difficulties in performing the iterations necessary for the modal estimators on the UNIVAC 1100 at the University of Wisconsin, Milwaukee. A total of 1440 modal-type estimates was calculated at no great expense. Other researchers have reported serious problems in implementing the iterative procedure for the m-group model (see e.g. Molenaar (1978)). It is likely that our good fortune is attributable to one or more of three possible causes: (1) using a prior that prevents the $\underline{\beta}$ estimators from being all equal, (2) using artificial data for which the hierarchical model was exactly true, and (3) restricting attention to the simple case $p = 2$.

5. CONCLUSION

In this paper we discuss the Bayesian analysis of a random coefficient m-group regression model using a more plausible stage III prior specification than has been previously considered. We show that in simple cases the resulting posterior distribution reduces to a form for which first- and second-order moments may be calculated, but in general the posterior distribution is not amenable to such analysis. We then consider point estimators of the linear parameters of the model and examine, via a Monte Carlo study, the impacts of a number of factors on the point estimators. In the simple model we consider, the O'Hagan estimator usually displays smaller mean squared error than the competing estimators.

APPENDIX A

Derivation of Modes of Wishart and Inverted Wishart Distributions

Properties of the Wishart and inverted Wishart distributions are discussed in many sources (see e.g. Anderson (1958), Press (1972), and Srivastava and Khatri (1979)). Since the formulas for their modes seem not to be well known, however, although implicit in the work of Lindley and Smith (1972) and O'Hagan (1976), we shall sketch their derivation here. We require two lemmas.

Lemma 1: If $\underset{\sim}{X}$ is symmetric and non-singular, then

$$d|\underset{\sim}{X}|/d\underset{\sim}{X} = |\underset{\sim}{X}|\{2\underset{\sim}{X}^{-1} - \text{diag}(\underset{\sim}{X}^{-1})\}.$$

Proof: The proof follows from an expansion of a determinant by cofactors and from the formula for the inverse of a matrix in terms of determinants of cofactors. ■

Lemma 2: If both $\underset{\sim}{B}$ and $\underset{\sim}{X}$ are symmetric, then

$$d(\text{tr}\underset{\sim}{B}\underset{\sim}{X})/d\underset{\sim}{X} = 2\underset{\sim}{B}\text{-diag}(\underset{\sim}{B}).$$

Proof:

$$\begin{aligned}\partial(\text{tr}\underset{\sim}{B}\underset{\sim}{X})/\partial X_{ij} &= \frac{\partial}{\partial X_{ij}}[\Sigma\Sigma b_{k\ell}X_{k\ell}] \\ &= \Sigma\Sigma b_{k\ell}(\partial X_{k\ell}/\partial X_{ij}) \\ &= \begin{cases} 2b_{ij} & \text{if } i \neq j \\ b_{ii} & \text{if } i = j \end{cases}\end{aligned}$$

■

Now to say that $\underset{\sim}{X} \sim W(\underset{\sim}{A}, \nu, p)$ means that the density $f(\underset{\sim}{X})$ is proportional to $|\underset{\sim}{X}|^{\frac{1}{2}(\nu-p-1)} \exp\{-\frac{1}{2}\mathrm{tr}\underset{\sim}{A}^{-1}\underset{\sim}{X}\}$. Then

$$\ell nf(\underset{\sim}{X}) = c + \tfrac{1}{2}(\nu-p-1)\ell n|\underset{\sim}{X}| - \tfrac{1}{2}\mathrm{tr}\underset{\sim}{A}^{-1}\underset{\sim}{X},$$

where c is a proportionality constant, and

$$d\ell nf(\underset{\sim}{X})/d\underset{\sim}{X} = \tfrac{1}{2}(\nu-p-1)|\underset{\sim}{X}|\{2\underset{\sim}{X}^{-1} - \mathrm{diag}(\underset{\sim}{X}^{-1})\}/|\underset{\sim}{X}|$$

$$- \tfrac{1}{2}\{2\underset{\sim}{A}^{-1} - \mathrm{diag}(\underset{\sim}{A}^{-1})\},$$

by lemmas 1 and 2. If this expression is equated to zero, then

$$2\{(\nu-p-1)\underset{\sim}{X}^{-1} - \underset{\sim}{A}^{-1}\} - \mathrm{diag}\{(\nu-p-1)\underset{\sim}{X}^{-1} - \underset{\sim}{A}^{-1}\} = \underset{\sim}{0},$$

which is possible only if $(\nu-p-1)\underset{\sim}{X}^{-1} - \underset{\sim}{A}^{-1} = \underset{\sim}{0}$, i.e., if $\underset{\sim}{X} = (\nu-p-1)\underset{\sim}{A}$.

To say $\underset{\sim}{Y} \sim W^{-1}(\underset{\sim}{A}^{-1}, \nu+p+1, p)$ means that $f(\underset{\sim}{Y})$ is proportional to $|\underset{\sim}{Y}|^{-\frac{1}{2}(\nu+p+1)} \exp\{-\frac{1}{2}\mathrm{tr}\underset{\sim}{A}^{-1}\underset{\sim}{Y}^{-1}\}$. The derivation of the mode $(\nu+p+1)^{-1}\underset{\sim}{A}^{-1}$ proceeds in exactly the same manner as for the Wishart density.

APPENDIX B

A Canonical Transformation

We sketch the steps leading from expression (6a) to expression (8). Let $\underline{H}$, $\underline{Q}_m$, $\underline{\Lambda}_m$, and $\underline{\Gamma}$ be defined as in the text above expression (7). Define the mp×1 vectors $\underline{\beta} = P(\underline{B}')$, $\underline{\delta} = P(\underline{\Gamma}')$, and $\hat{\underline{\beta}} = P(\underline{\hat{B}}')$, where $P(\cdot)$ is the pack operator. If $\hat{\underline{\Gamma}}$ is defined by $\hat{\underline{B}} = m^{-\frac{1}{2}}\underline{H}\hat{\underline{\Gamma}}\underline{U}$, then we can define $\hat{\underline{\gamma}} = P(\hat{\underline{\Gamma}}')$. With these definitions

$$(\underline{\beta}-\hat{\underline{\beta}}) = m^{-\frac{1}{2}}(\underline{H}\otimes\underline{U}')(\underline{\gamma}-\hat{\underline{\gamma}}),$$

since $\underline{Y} = \underline{A}\underline{X}$ implies $P(\underline{Y}) = (\underline{I}\otimes\underline{A})P(\underline{X})$ and $\underline{Y} = \underline{X}\underline{B}$ implies $P(\underline{Y}) = (\underline{B}'\otimes\underline{I})P(\underline{X})$. Then

$$\sum_{i=1}^{m}(\underline{\beta}_i-\hat{\underline{\beta}}_i)'\underline{X}_i'\underline{X}_i(\underline{\beta}_i-\hat{\underline{\beta}}_i)$$

$$= (\underline{\beta}-\hat{\underline{\beta}})'\begin{bmatrix}\underline{X}_1'\underline{X}_1 & \underline{0} & \cdots & \underline{0}\\ \underline{0} & \underline{X}_2'\underline{X}_2 & \cdots & \underline{0}\\ \underline{0} & \underline{0} & \cdots & \underline{X}_m'\underline{X}_m\end{bmatrix}(\underline{\beta}-\hat{\underline{\beta}})$$

$$= (\underline{\gamma}-\hat{\underline{\gamma}})'(\underline{H}'\otimes\underline{U})m^{-1}\begin{bmatrix}\underline{X}'\underline{X}_1 & \underline{0} & \cdots & \underline{0}\\ \underline{0} & \underline{X}_2'\underline{X}_2 & \cdots & \underline{0}\\ \underline{0} & \underline{0} & & \underline{X}_m'\underline{X}_m\end{bmatrix}(\underline{H}\otimes\underline{U}')(\underline{\gamma}-\hat{\underline{\gamma}})$$

$$= r(\underline{\gamma}-\hat{\underline{\gamma}})'\nu_1^{-1}\underline{S}(\underline{\gamma}-\hat{\underline{\gamma}}),$$

where $\underline{S}$ and ν_1 are defined as in the text of the paper.

REFERENCES

[1] Anderson, T.W. (1958), An Introduction to Multivariate Statistical Analysis. New York: Wiley.

[2] Fortney, W.G. and Grupe, M.R. (1976). Contours of the bivariate Wishart distribution. Technical Report No. 452, Department of Statistics, University of Wisconsin, Madison.

[3] Fortney, W.G. and Miller, R.B. (1980a). Practical examples of random coefficient regression analysis from a Bayesian point of view. Technical Report, Department of Statistics, University of Wisconsin, Madison, (forthcoming).

[4] Fortney, W.G. and Miller, R.B. (1980b). Bayesian analysis in random coefficient multiple regression. Technical Report, Department of Statistics, University of Wisconsin, Madison, (forthcoming).

[5] Goel, P.K. and DeGroot, M.H. (1979). Information about hyperparameters in hierarchical models. Technical Report No. 160, Department of Statistics, Carnegie-Mellon University.

[6] Lindley, D.V. and Smith, A.F.M. (1972). "Bayes estimates for the linear model". J.R. Statist. Soc. B, 34, 1-41.

[7] Molenaar, I.W. (1978). A fast solution of the Lindley equations for the m-group regression problem. Technical Report No. 78-3, Iowa Testing Programs, University of Iowa.

[8] Novick, M.R., Jackson, P.H., Thayer, D.T., and Cole, N. (1972). "Estimating multiple regressions in m-groups: a cross-validation study". British J. Math. and Statist. Psych. 25, 33-50.

[9] O'Hagan, A. (1976). "On posterior joint and marginal modes". Biometrika 63, 329-333.

[10] Press, S.J. (1972). Applied Multivariate Analysis. New York: Holt, Rinehart, and Winston.

[11] Srivastava, M.S. and Khatri, C.G. (1979). An Introduction to Multivariate Statistics. New York: North Holland.

[12] Tiao, G.C. and Zellner, A. (1964). "Bayes's theorem and the use of prior knowledge in regression analysis. Biometrika 51, 219-230.

Table 1. Definition of Cases for Monte Carlo Study

n_i	m	r_x	ρ	Case
50	8	0.0	0.0	A_1
50	8	0.0	0.5	B_1
50	8	0.5	0.0	C_1
50	8	0.5	0.5	D_1
50	20	0.0	0.0	E_1
50	20	0.0	0.5	F_1
50	20	0.5	0.0	G_1
50	20	0.5	0.5	H_1
100	8	0.0	0.0	A_2
100	8	0.0	0.5	B_2
100	8	0.5	0.0	C_2
100	8	0.5	0.5	D_2
100	20	0.0	0.0	E_2
100	20	0.0	0.5	F_2
100	20	0.5	0.0	G_2
100	20	0.5	0.5	H_2

Table 2. Average value of d_2 over 30 replications of each case.

Values for joint modal, marginal modal, and O'Hagan estimate (percentage of ordinary least squares in parentheses). Ranks are of the numbers in parentheses for the best estimates for each case.

$n_i = 50$

	$r_x = 0$		$r_x = .5$	
m	$\rho = 0$	$\rho = 0.5$	$\rho = 0$	$\rho = 0.5$
8	Case A_1	Case B_1	Case C_1	Case D_1
	3.740(97.8)	4.563(92.7)	4.648(92.5)	7.891(95.4)
	3.725(97.4)*	4.594(93.3)	4.670(92.9)	7.683(92.9)*
	3.734(97.7)	4.558(92.6)*	4.638(92.3)*	7.746(93.7)
	Rank 16	Rank 7	Rank 6	Rank 8.5
20	Case E_1	Case F_1	Case G_1	Case H_1
	3.501(95.6)	4.592(90.5)	5.136(93.2)	7.246(84.4)
	3.496(95.5)*	4.594(90.5)	5.125(93.0)	7.183(83.6)
	3.496(95.5)*	4.576(90.1)*	5.117(92.9)*	7.130(83.0)*
	Rank 14	Rank 4	Rank 8.5	Rank 1

$n_i = 100$

	$r_x = 0$		$r_x = .5$	
m	$\rho = 0$	$\rho = 0$	$\rho = 0$	$\rho = 0$
8	Case A_2	Case B_2	Case C_2	Case D_2
	3.730(96.0)	4.843(94.0)	4.492(95.5)	7.300(85.6)
	3.740(96.2)	4.850(94.2)	4.465(94.9)	7.358(86.3)
	3.727(95.9)*	4.823(93.7)*	4.469(95.0)*	7.268(85.3)*
	Rank 15	Rank 12	Rank 13	Rank 2
20	Case E_2	Case F_2	Case G_2	Case H_2
	3.583(93.3)	4.226(91.4)	5.200(93.8)	7.913(92.8)
	3.589(93.5)	4.215(91.1)	5.177(93.4)	7.663(89.9)
	3.581(93.3)*	4.204(90.9)*	5.165(93.2)*	7.494(87.9)*
	Rank 11	Rank 5	Rank 10	Rank 3

Table 3. Average d_2 Over Various Cases

Marginal Average d_2 (240 simulations)

	Joint Mode	Marginal Mode	O'Hagan
$n_i = 50$	5.165	5.134	5.124
$n_i = 100$	5.161	5.132	5.091
Difference	.004	.002	.033
$m = 8$	5.151	5.136	5.120
$m = 20$	5.137	5.130	5.095
Difference	-.024	.006	.025
$r_x = 0$	4.097	4.100	4.087
$r_x = .5$	6.228	6.166	6.128
Difference	-2.131	-2.066	-2.041
$\rho = 0$	4.254	4.248	4.241
$\rho = .5$	6.072	6.018	5.975
Difference	-1.818	-1.770	-1.734

Overall Average d_2 (480 simulations)

Joint Mode	Marginal Mode	O'Hagan
5.163	5.113	5.108